



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **10289247 A**

(43) Date of publication of application: 27 . 10 . 98

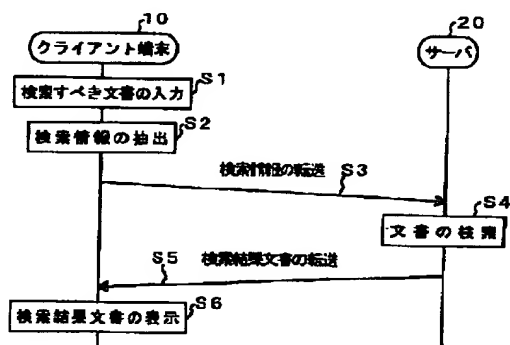
(51) Int. Cl

G06F 17/30(21) Application number: **09097833**(22) Date of filing: **16 . 04 . 97**(71) Applicant: **NIPPON TELEGR & TELEPH
CORP <NTT>**(72) Inventor: **OGURA KENJI
OGURO MASAMI
NAKA MURA OSAMU
AKIYAMA TERUO****(54) DOCUMENT RETRIEVING METHOD AND
SYSTEM****(57) Abstract:**

PROBLEM TO BE SOLVED: To provide a document retrieving method and system which retrieve even when a file name and a keyword are not clear and also retrieve with non-character information in retrieving mass document.

SOLUTION: A client terminal 10 partially or totally inputs document to be retrieved (S1), retrieval information is extracted from inputted document information (S2), the retrieval information is transferred to a server 20 through a communication line (S3), the server 20 side retrieves document that is previously accumulated based on the received retrieval information (S4) and transfers it to the terminal 10 (S5), and the terminal 10 shows the document of a retrieval result (S6).

COPYRIGHT: (C)1998,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-289247

(43) 公開日 平成10年(1998)10月27日

(51) Int.Cl.⁵
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/40 3 1 0 F
3 7 0 A
15/403 3 1 0 C

審査請求 未請求 請求項の数 5 O L (全 8 頁)

(21) 出願番号 特願平9-97833

(22) 出願日 平成9年(1997)4月16日

(71) 出願人 000004226
日本電信電話株式会社
東京都新宿区西新宿三丁目19番2号
(72) 発明者 小倉 健司
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(72) 発明者 小黑 雅己
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(72) 発明者 中村 修
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(74) 代理人 弁理士 小笠原 吉義 (外1名)
最終頁に続く

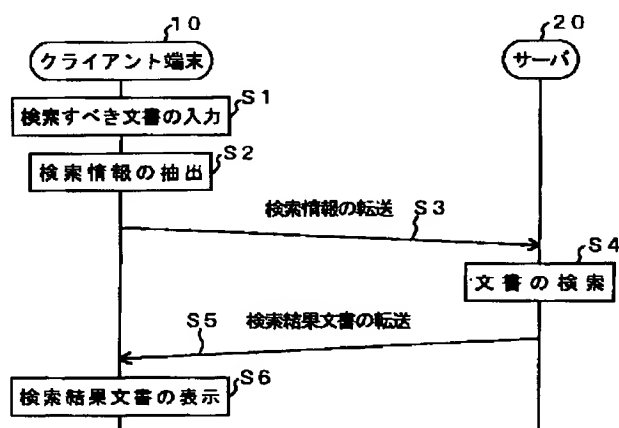
(54) 【発明の名称】 ドキュメント検索方法およびシステム

(57) 【要約】

【課題】 大量のドキュメントを検索する場合に、ファイル名やキーワードが不明であっても検索が可能であり、また非文字情報による検索が可能であるドキュメント検索方法およびシステムを実現する。

【解決手段】 クライアント端末10で検索すべき文書の一部あるいは全てを入力し(S1)、入力された文書情報から検索情報を抽出し(S2)、通信回線を介して検索情報をサーバに転送し(S3)、サーバ20側では受信した検索情報に基づき予め蓄積している文書を検索して(S4)、これをクライアント端末10に転送し(S5)、クライアント端末10で検索結果の文書を表示する(S6)。

本発明の原理説明図



【特許請求の範囲】

【請求項1】 サーバ/クライアントシステムにおいて、サーバに蓄積された文書情報をクライアント端末からの指示により検索するドキュメント検索方法であって、クライアント端末から検索すべき文書の一部あるいは全てを入力し、入力された文書情報から検索情報を抽出し、サーバに蓄積された複数の文書情報の中から、抽出された検索情報の内容と一致する内容を含む文書を検索し、該検索された文書情報をクライアント端末に表示することを特徴とするドキュメント検索方法。

【請求項2】 クライアント端末から検索すべき文書の一部あるいは全てをビットマップ情報として入力し、該ビットマップ情報から文字部分を認識して文字コード列に変換し、文字コードでサーバに蓄積された複数の文書情報の中から一致する文字コード列を含む文書情報を検索・表示することを特徴とする請求項1記載のドキュメント検索方法。

【請求項3】 クライアント端末から検索すべき文書の一部あるいは全てをビットマップ情報として入力し、サーバに蓄積された文書情報の中から該ビットマップ情報の一部あるいは全てが相似的に一致するイメージ情報を含む文書情報を検索・表示することを特徴とする請求項1記載のドキュメント検索方法。

【請求項4】 検索すべき文書情報を指定するクライアント端末と、複数の文書情報を蓄積し、クライアント端末からの指示により特定の文書情報を検索するサーバとを有するドキュメント検索システムであって、前記クライアント端末は、検索すべき文書情報の一部あるいは全てをビットマップ情報として入力する文書入力手段と、該ビットマップ情報から文字部分を切り出し、文字コードに変換する文字認識手段と、文字認識された文字コード列を蓄積する検索情報蓄積手段と、該検索情報を前記サーバに転送する転送手段と、前記サーバから転送される検索結果としての文書情報を表示する文書表示手段とを有し、前記サーバは、複数の文書情報を蓄積する文書蓄積手段と、前記クライアント端末からの検索情報を受信する受信手段と、受信した検索情報である文字コード列を蓄積する検索用文字コード列蓄積手段と、前記文書蓄積手段に蓄積された複数の文書情報の文字コード部分と前記検索用文字コード列蓄積手段に蓄積された文字コードとの一致をとり、その一致度合を判定する一致判定手段と、前記一致判定手段の出力により、前記文書蓄積手段に蓄積された複数文書情報の中から特定の文書情報のファイル名を蓄積するファイル名蓄積手段と、前記文書蓄積手段に蓄積された全文書の一致判定が終了した時点で、前記ファイル名蓄積手段に蓄積されたファイル名で特定される文書情報を前記クライアント端末に転送する転送手段とを有することを特徴とするドキュメント検索システム。

【請求項5】 検索すべき文書情報を指定するクライ

アント端末と、複数の文書情報を蓄積し、クライアント端末からの指示により特定の文書情報を検索するサーバとを有するドキュメント検索システムであって、前記クライアント端末は、検索すべき文書情報の一部あるいは全てをビットマップ情報として入力する文書入力手段と、該ビットマップ情報から特定の領域を検索情報として切り出す領域切り出し手段と、切り出された検索情報であるビットマップ情報を蓄積する検索情報蓄積手段と、該検索情報を前記サーバに転送する転送手段と、前記サーバから転送される検索結果としての文書情報を表示する文書表示手段とを有し、前記サーバは、複数の文書情報を蓄積する文書蓄積手段と、前記クライアント端末からの検索情報を受信する受信手段と、受信した検索情報であるビットマップ情報を蓄積する検索用ビットマップ情報蓄積手段と、前記文書蓄積手段に蓄積された複数の文書情報の中からビットマップ情報の一部を切り出し、該切り出されたビットマップ情報に拡大縮小などの変換を行い、前記検索用ビットマップ情報蓄積手段に蓄積されたビットマップ情報との一致をとり、その一致度合を判定する一致判定手段と、前記一致判定手段の出力により、前記文書蓄積手段に蓄積された複数文書情報の中から特定の文書情報のファイル名を蓄積するファイル名蓄積手段と、前記文書蓄積手段に蓄積された全文書の一致判定が終了した時点で、前記ファイル名蓄積手段に蓄積されたファイル名で特定される文書情報を前記クライアント端末に転送する転送手段とを有することを特徴とするドキュメント検索システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ドキュメント検索方法およびシステムに係わり、特にクライアント端末から文書の一部あるいは全てを入力し、入力情報から検索情報を生成して、これに基づきサーバにおいて当該文書を検索するドキュメント検索方法およびシステムに関する。

【0002】

【従来の技術】従来のドキュメント検索システムでは、文書単位にファイル名を付与してデータベースなどに蓄積し、検索時には予め付与されたファイル名を指定したり、文書の中に含まれる特定のキーワードを指定して、予め設定しておいたキーワードとの照合あるいは全文検索により当該文書を検索している。

【0003】

【発明が解決しようとする課題】しかしながら、上記の従来方法では、ファイル名、キーワードなどに文字数の制限があったりして、ファイル名あるいはキーワードで検索しようとした場合、自分で付与したにも関わらず、ファイル名あるいはキーワードを忘れてしまったり、さらに、他人が付与したファイル名あるいはキーワードはほとんど分からないといった問題が生じていた。

【0004】また、ある任意の文字列による全文検索の場合には、クライアント端末からキーボードなどにより文字列を入力しなければならないというような問題や、文書に含まれる図形などの非文字情報からの検索は不可能であるというような問題があった。

【0005】本発明は、上記の点に鑑みてなされたもので、紙ベースで保管していた文書あるいは記憶に頼った手書き文書などを入力するだけで目的とする文書の検索を可能とするドキュメント検索方法およびシステムを提供することを目的とする。

【0006】

【課題を解決するための手段】図1は、本発明の原理を説明するための図である。本発明は、クライアント端末10において検索すべき文書の一部あるいは全てを入力し（ステップS1）、入力された文書情報から検索情報を抽出し（ステップS2）、通信回線を介して検索情報をサーバ20に転送し（ステップS3）、サーバ20側では受信した検索情報に基づき文書を検索して（ステップS4）、これをクライアント端末10に転送し（ステップS5）、クライアント端末10でサーバ20側から送られた検索結果文書を表示する（ステップS6）ものである。

【0007】図2は、本発明の原理構成図である。クライアント端末10は、文書入力手段11と検索情報抽出手段12とを持つ。文書入力手段11は、検索すべき文書の一部あるいは全てを入力するものである。検索情報抽出手段12は、入力された文書情報から検索情報を抽出するものである。

【0008】サーバ20は、検索情報蓄積手段21と文書蓄積手段22と一致判定手段23とを持つ。検索情報蓄積手段21は、クライアント端末10から受信した検索情報を蓄積するものである。文書蓄積手段22は、検索対象となる複数の文書を蓄積し保管するものである。一致判定手段23は、複数の文書情報の文字コード部分（またはビットマップ情報）と、検索情報の文字コード部分（またはビットマップ情報）との一致をとり、その一致度合いを判定するものである。

【0009】

【発明の実施の形態】以下に、本発明の実施の一形態を説明する。図3および図4は、本発明のシステム構成図であって、特に、図3はクライアント端末のシステム構成例、図4はサーバのシステム構成例を示す。

【0010】図3および図4に示すように、本ドキュメント検索システムは、通信網30を介してクライアント端末10およびサーバ20が接続される構成となる。クライアント端末10は、図3に示すように、CPU・メモリ110、文書情報をビットマップ情報として入力するカメラあるいはファクシミリ装置などの文書入力装置120、CPU・メモリ110の中にあつて、文書入力装置120から入力されたビットマップ情報から文字を

認識する文字認識機構111、同じくCPU・メモリ110の中にあつて文字認識機構111で認識された文字コード列を蓄積する検索情報蓄積機構112、該検索情報を通信網30に送り出し、サーバ20に転送する通信制御装置130、およびサーバ20から転送される検索結果の文書情報を表示する表示装置140から構成される。

【0011】また、サーバ20は、図4に示すように、CPU・メモリ210、複数の文書情報を蓄積する文書蓄積装置220、CPU・メモリ210の中にあつてクライアント端末10から転送された検索情報である文字コード列を蓄積する検索用文字コード列蓄積機構211、同じくCPU・メモリ210の中にあつて該検索情報と文書蓄積装置220に蓄積されている文書情報との一致を判定する一致判定機構212、同じくCPU・メモリ210の中にあつて一致判定機構212で検索された文書情報のファイル名を蓄積するファイル名蓄積機構213、およびクライアント端末10から検索情報を受信したり、ファイル名蓄積機構213に蓄積されたファイル名で特定される文書情報をクライアント端末10に転送する通信制御装置230から構成される。

【0012】上記の構成の動作を説明する。図5（A）は、本発明のクライアント端末10における処理の概要を示すフローチャートである。

【0013】ステップS11では、検索すべき文書の写しなどを文書入力装置120からビットマップ情報として入力する。ステップS12では、文書入力装置120から入力されたビットマップ情報に対し、CPU・メモリ110内の文字認識機構111にてノイズ除去、傾き補正などのイメージ整形処理を前処理とする文字認識を行う。文字認識の具体的処理方法については、例えば萩田他の「外部方向寄与度特徴による漢字の識別」（電子情報通信学会論文誌Vol. J66-D, No. 10, 1983）に示される技術を用いることにより実現する。一般に用いられている文字認識技術を用いてもよい。

【0014】ステップS13では、文字認識機構111で認識された結果である文字コードを検索情報蓄積機構112に順次格納する。ステップS14では、入力されたビットマップ情報に含まれる全ての文字に対する処理が終了したかを判断し、終了していなければステップS12～S14を繰り返す。

【0015】ステップS15では、入力されたビットマップ情報に含まれる全ての文字に対する処理が終了したら、検索情報蓄積機構112に蓄積された検索情報を通信制御装置130よりサーバ20に転送する。

【0016】ステップS16では、サーバ20からの検索結果の受信を待ち、検索結果の文書を受信したら、ステップS17において検索文書を表示装置140に表示する。

【0017】図5（B）は、本発明のサーバ20にお

る処理の概要を示すフローチャートである。ステップS21では、通信制御装置230で、クライアント端末10から検索情報を受信し、これをCPU・メモリ210内の検索用文字コード列蓄積機構211に格納する。

【0018】ステップS22では、複数の文書情報を蓄積している文書蓄積装置220から、一つの文書情報をCPU・メモリ210に読み出す。ステップS23では、読み出された文書に対し、検索用文字コード列蓄積機構211に蓄積された検索情報に基づき一致判定機構212にて一致の判定処理を行う。一致を判定する具体的処理方法については、例えば宮原他の「SIMD型並列プロセッサを用いたフルテキスト検索」(情報処理学会論文誌Vol.33 No.3, 1992)に示される技術を用いることにより実現する。

【0019】この「SIMD型並列プロセッサを用いたフルテキスト検索」の技術は、SIMD型2次元アレイプロセッサとホストコンピュータ(パソコン)によって、高速のフルテキスト検索機能を実現したものであり、パソコンから検索プログラムと文書データとをロードしておき、必要に応じて検索単語と検索条件とを入力することで検索を行うことができるようにしたものである。文書データの水平格納によるビットシリアル型の処理方式を採用する。部分一致検索では、異字許容照合や単語内ワイルドカード照合が、それぞれ全文照合や絞り込み照合とほぼ同程度の速度で検索できる。もちろん、ステップS23における一致判定処理では、この方法に限らず、通常の単一プロセッサによる検索手段等を用いてもよい。

【0020】ステップS24で一致の判定処理結果をチェックし、一致しない場合には、ステップS22の処理へ戻って、次の文書に対する処理を繰り返す。ステップS25では、一致判定機構212にて一致が判定された場合に、その文書のファイル名をファイル名蓄積機構213に格納する。

【0021】ステップS26では、検索結果であるファイル名蓄積機構213に格納されているファイル名で特定される文書情報を、通信制御装置230からクライアント端末10に転送する。

【0022】上述のように、本実施の形態では、ビットマップ情報から抽出される文字コードを用いて文書を検索する例を説明したが、入力したビットマップ情報そのものを切り出し、これをキーとして図形処理、画像処理技術により被検索情報との一致をとり、該当する文書情報を検索することも可能である。このとき、必ずしも大きさまでも一致する必要はなく、相似的に一致すればよい。このため、必要であれば、検索情報であるビットマップ情報に拡大または縮小などの変換を行い、一致を調べる。入力したビットマップ情報から特定の領域を検索情報として切り出す場合には、例えば入力したビットマップ情報を表示装置に表示し、その中で利用者に会話的

に領域を指定させるなどの方法を用いることができる。

【0023】また、本実施の形態では、検索される文書が一つとして説明しており、文字認識精度については言及していないが、実際には入力された文字イメージが100%認識できることは稀であるため、ある程度の認識精度を確保できれば、複数の候補文書を提示して利用者に判断させることでよい。例えば、認識すべき文字数の半分以上でも認識できれば、その認識できた部分で検索した文書を、候補文書として利用者に提示し、複数の候補文書の中から利用者が詳細を知りたい文書情報を利用者に選択させるようにしてもよい。

【0024】本発明は、以上説明した例に限定されることなく、特許請求の範囲における各請求項に記載した技術の範囲内で種々変更・応用が可能であることは言うまでもない。

【0025】

【発明の効果】上述のように、本発明のドキュメント検索方法およびシステムによれば、蓄積されている大量の文書情報のファイル名、キーワードが不明であったり、忘れていても、紙ベースの文書の写しの一部でもあれば、蓄積されている電子データを検索できるばかりか、従来のフルテキストサーチによる検索手段では不可能であった図形をもキーとした検索が可能となる。

【図面の簡単な説明】

【図1】本発明の原理説明図である。

【図2】本発明の原理構成図である。

【図3】本発明のクライアント端末のシステム構成例を示す図である。

【図4】本発明のサーバのシステム構成例を示す図である。

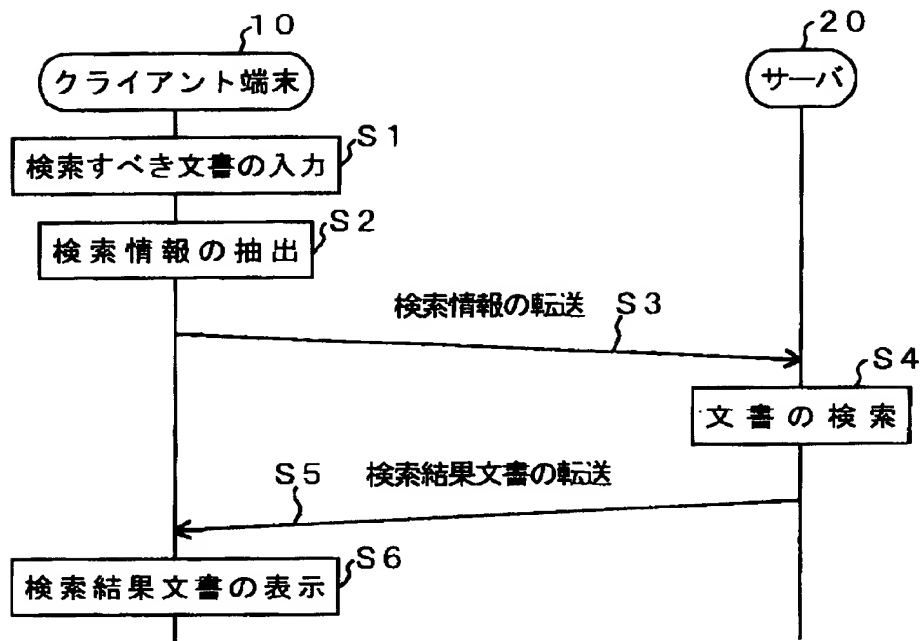
【図5】本発明のクライアント端末およびサーバの処理の概要を示すフローチャートである。

【符号の説明】

- 10 クライアント端末
- 110 (クライアント端末内の) CPU・メモリ
- 111 (CPU・メモリ内の) 文字認識機構
- 112 (CPU・メモリ内の) 検索情報蓄積機構
- 120 (クライアント端末内の) 文書入力装置
- 130 (クライアント端末内の) 通信制御装置
- 140 (クライアント端末内の) 表示装置
- 20 サーバ
- 210 (サーバ内の) CPU・メモリ
- 211 (CPU・メモリ内の) 検索用文字コード列蓄積機構
- 212 (CPU・メモリ内の) 一致判定機構
- 213 (CPU・メモリ内の) ファイル名蓄積機構
- 220 (サーバ内の) 文書蓄積装置
- 230 (サーバ内の) 通信制御装置
- 30 通信網

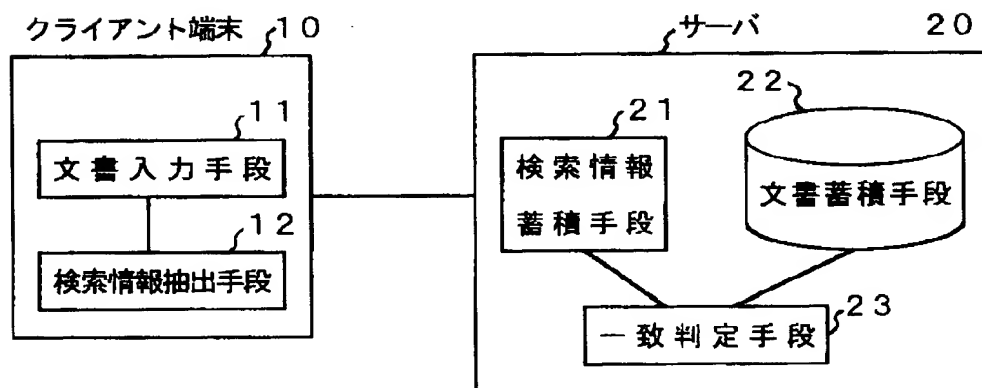
【図1】

本発明の原理説明図



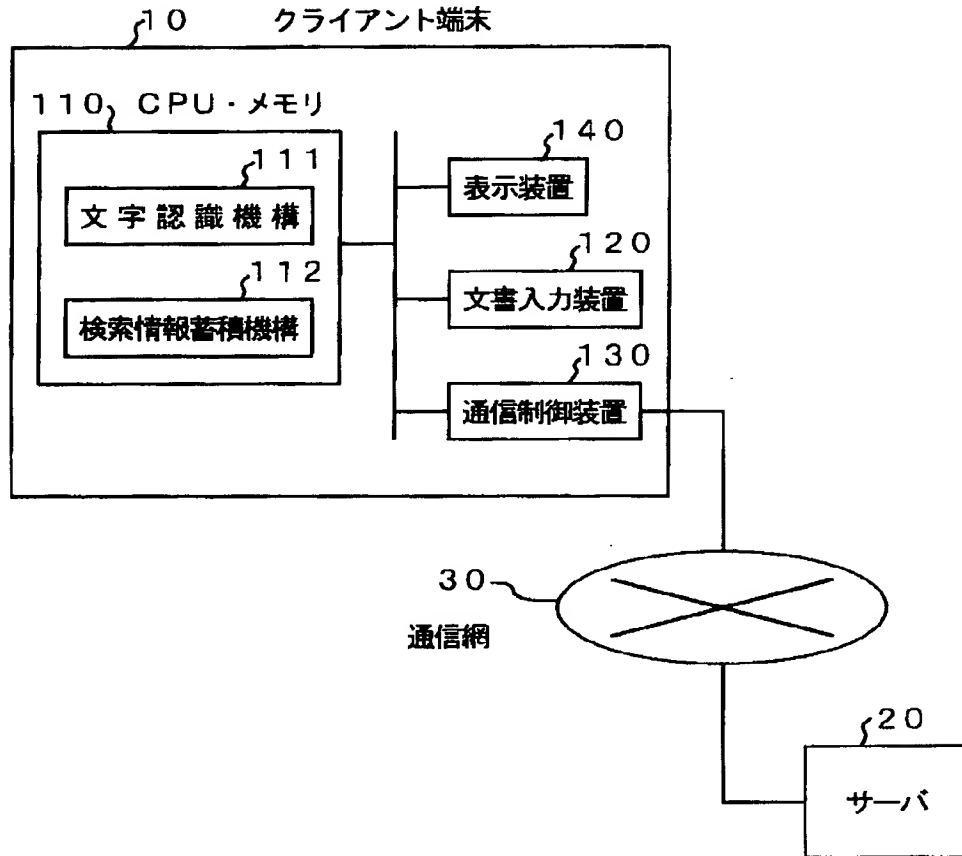
【図2】

本発明の原理構成図



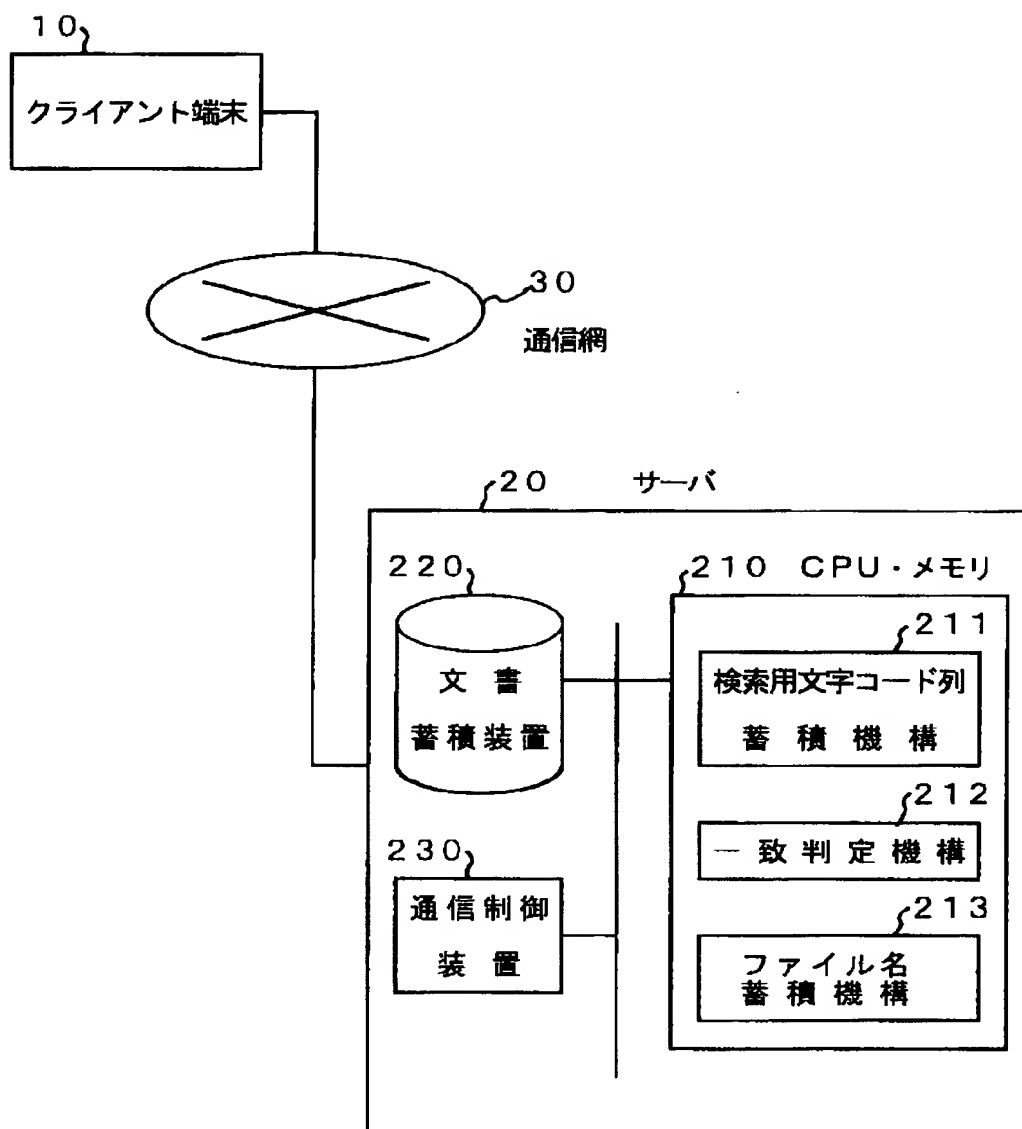
【図3】

クライアント端末のシステム構成例



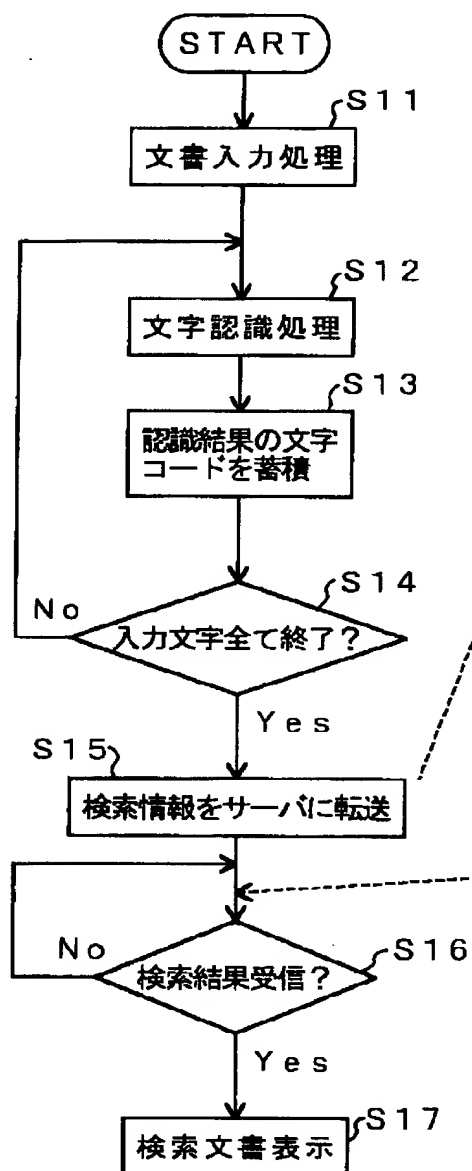
【図4】

サーバのシステム構成例

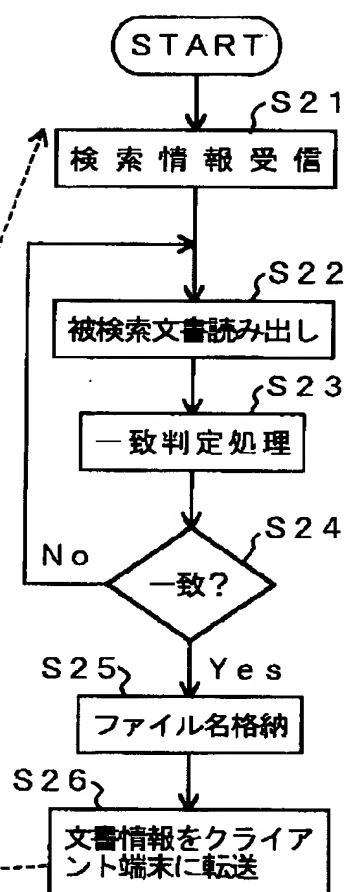


【図5】

(A) クライアント端末の処理フロー



(B) サーバの処理フロー



フロントページの続き

(72)発明者 秋山 照雄
 東京都新宿区西新宿三丁目19番2号 日本
 電信電話株式会社内